THE USE OF WALKING SOUNDS IN SUPPORTING AWARENESS

Kaj Mäkelä, Jaakko Hakulinen & Markku Turunen

Department of Computer and Information Sciences
33014 University of Tampere
FINLAND
{kaj.makela, jaakko.hakulinen, markku.turunen}@cs.uta.fi

ABSTRACT

There is need for applications that keep the users aware of various things happening around them. One of such things is the presence of co-workers in office environments. Playing people's walking sounds in the background as cues of their presence came up as one way to provide such information. The walking sounds have very favorable properties; as a part of the normal sound ecology in an office environment they are non-obtrusive. In addition, they have a natural association to the person they present. To study how well people can identify recorded walking sounds of their co-workers we organized a listening experiment. Without any learning the identification rates were 13%, however after short teaching session rates increased to 66%. The results show that walking sounds do contain enough information to convey the identity of the person but careful design of the sounds is necessary. In this paper we also present how walking sounds can be used as a part of a ubiquitous computing system.

1. INTRODUCTION

As computing systems become more pervasive and ubiquitous, we can build applications that help users to maintain their awareness of things happening around them. Sound provides an excellent tool for such systems, as it is able to provide information in the background without a need for a user to attend to it all the time. One specific application we have considered is keeping people aware of their co-workers in the workplace. When the users have such awareness, they can easily contact people face to face, which is likely to support also informal communication.

Providing such information with audio is not however trivial. Any solution is a compromise between two parameters: how well the information is communicated to the users and how much annoyance and obtrusion the system causes. Considering these parameters, we came up with an idea of using walking sounds as a part of our ubiquitous computing system *Doorman [6]*.

Walking sounds are naturally occurring non-obtrusive sound cues in office environments; they are natural part of the sound ecology of the office. The walking sounds do not usually distract people from their work. However walking sounds do indicate the presence of the person and it is normal for people to react to the walking sounds of somebody who theyíve been expecting. People walking by an office in a hallway make their own, recognizable sounds. These walking sounds are not very easily recognized, but our informal finding is that the people who often interact or meet may learn to recognize each otherís walking sounds without explicit training. In our design the walking sound is used as a metaphor for the presence of a person. We consider the metaphor intuitive and intelligible, because the walking sound normally occurs when the person is in the vicinity. The association for the person already exists.

The human recognition of walking sounds has been studied and they have also been used in auditory interfaces. Li et al. [1] conducted experiments, which showed that the listeners are able to recognize gender from the properties of walking sounds.

In computer systems walking sounds have been used for example as auditory icons in Europarc's RAVE system [2] where they provided peripheral awareness that someone is watching through a web cam. The used walking sound was generic and it was not used for the identification of the watcher.

The recorded walking sounds do differ from the naturally occurring walking sounds and several things need to be considered when such sounds are recorded and played.

To ensure that the sound is perceivable, it should be played with adequate volume. The sound should be in a relation to the current sound ecology, so that it would not be masked by other sounds and in return, would not mask other important sounds. The presentation of the recorded walking sound should be adjusted to the environment and the situation. The volume and spectral structure of the sound should be tested in the environment and adjusted correspondingly.

The counterpart to the demand for perceptivity is the demand for non-obtrusiveness. The sound must not disturb and interfere with the tasks and interaction of the people that are exposed to the sound stimuli. Non-obtrusiveness is enabled by using the sounds in a subtle manner, in the background. This means that the auditory cues should, for example emerge gradually and sudden changes should be avoided.

When sound cues are used as a secondary source of information, providing peripheral awareness, the system does not need to provide information every moment. The person needing the exact information can check the presence of the staff member in other ways.

The open questions regarding the use of walking sounds as the cues of presence of co-workers are twofold. First, it is not sure if the walking sounds are indeed non-obtrusive. The second question is whether people can identify the walking sounds with good enough precision for the system envisioned to be of any advantage. We consider that the second question is more crucial and therefore we arranged a listening test to see if walking sounds can be identified at any reasonable rate.

The rest of the paper describes the experiment starting with general discussion, and including description how the used walking sounds were recorded, the subjects, the experiment set up, the results and the discussion. Finally, we present how the walking sounds can be used as a part of a ubiquitous computing system.

2. Listening Experiment

The aim of the test was to evaluate the ability of people to identify the walking sounds of their co-workers on the base of their everyday experience. This was tested by playing recorded walking sounds to the subjects and asking their opinion on the identity of the walker.

Before the actual test we conducted a preliminary study. Three major issues were found: the recorded sounds need to be longer than a couple of seconds, they need to be carefully recorded and teaching is probably needed.

2.1. Subjects

The experiment was conducted on a group of people working closely together to find out if they were able to recognize or be able to learn to recognize each otherís walking sounds. An academic research group of thirteen persons working in the same premises was selected. The group consisted of nine men and four women between the ages 24 and 46. One person's ability to hear high frequencies had lowered and he had tinnitus, all the other group members had normal hearing.

2.2. Walking sounds

The walking sounds of each member were recorded in a laboratory environment. To gain samples that were possible to loop to continual walking sound and to minimize changes in the sound, an omni-directional microphone (AKG CK 32) was placed at the center of the room, pointing to the floor at forty centimeters height and each person was asked to walk circle with diameter of 2 meters around the microphone as naturally as possible. During the recording, the people wore their normal, everyday shoes and clothes.

Each person was recorded for thirty seconds, of which a suitable sample was normalized and trimmed to form a loop of five step pairs and duration of five to eight seconds. The samples contained the sounds of footsteps and other sounds occurring during walking, such as noises of clothes and accessories. People were photographed in the situation to document their appearance, current clothing and shoes.

2.3. Experiment Setup

The recognition of the sound samples was tested on twelve people of those thirteen whose walking was recorded. The testing sessions were held in normal meeting room that did not have soundproofing. The data projector used for the test was also making noise in the same room. The experiment was held in four sessions with one to eight subjects in each. They were asked not to communicate with each other during the test. The sounds were played through speakers (Genelec 1029A) and the distance to the listeners was approximately 2 meters.

In the test, the subjects were given four paged forms. Before the test listeners were asked if their hearing ability was normal. They were also asked how much they work together with the other group members on the scale from one to five. Furthermore the subjects were asked to literally describe each otherís walking sounds and tell if the description was based on knowledge or an impression. For both of the two listening tests in the experiment there was an evaluation form on its own page. For each sound sample, the subjects were asked identify the group member walking and put a number from one to five to report how confident they were about the identification. After the test the listeners were asked questions about the test, the parameters of walking sound they used for identification, and the design of sound samples.

To familiarize the listeners with the walking sounds before the test, four thirty-second-long walking sound samples of people outside the group were played. The listening experiments were held in three consecutively parts without breaks. The first part was a listening test where the people were played walking sound samples to be identified. There were thirteen different sounds and listeners were played thirty sound samples, so each sound was played from two to three times. Each sample was played for twenty seconds and there was ten second break between the samples. The second part was a teaching part, where the people were shown the picture of each person simultaneously with the same sound played as in the previous part. Each picture and sound was displayed for twenty seconds and there was a five second break between the displays. This was made two times consecutively. The third part of the test was similar to the first part and it tested the identification after the learning. The tests were conducted by using time controlled PowerPoint presentations. Each part was its own presentation containing needed instructions, timing, sounds and pictures. Each listening part lasted for fifteen minutes and teaching part took twelve minutes to run.

3. EXPERIMENT RESULTS

The results showed that the people were able to learn to recognize walking sounds. Teaching affected significantly ($t(11)=13.9$, $p<0.001$) the identification results.

In the first part of the experiment, the people were asked to identify walking sounds on the base of their current knowledge. The average identification rate per listener was 12.5 percent and rates altered between 3 and 20 percent. The average recognition rate of the listenersí own walking sounds was 28.6 percent. The average confidence to all identifications on the scale from 1 to 5 was 2.5. The average confidence given for the correct answers was 2.8, whereas the average confidence for in-correct answers was 2.5. The difference between the confidences was not statistically significant.

In the second part conducted after teaching part the average for correct identifications per listener was 66 percent, and the rate altered between 43 and 97 percent. The average recognition rate of the listenersí own walking sounds was 78.6 percent. The average confidence for all the answers was 4.0, for correct answers it was 4.3 and for incorrect answers 3.4. The difference between confidences of correct and incorrect answers was significant (t(11)=4.58, p<0.01). See table 1 for more detailed results.

	Correct		Confidence		
	Num (of 30)	%	All	Correct	Incorrect
Average	3,8 ➔ 19,8	13 ➔ 66	2,5➔4,0	2,8➔4,3	2,5➔3,4
max	6 ➔ 29	20 ➔ 97	3,8➔4,7	4,0➔4,9	3,9➔5,0
min	1 ➔ 13	3 ➔ 43	1,2➔3,1	1,5➔3,5	1,1➔1,7
std. dev.	2,0 ➔ 5,2	6,7➔17,2	0,9➔0,5	0,9➔0,4	0,9➔1,0

Table 1. *The walking sound identification results before (left values) and after (right values) the teaching session.*

There was no correlation between the listener's own evaluation on the degree of the co-operation or connection with the person and the recognition rate of the person's walking sounds.

4. DISCUSSION

The experiment showed that the users were able to identify the walking sounds. However, attaining the adequate level of identification required that the walking sounds were taught to the users. The experiment brought up some issues on the design, teaching and identification of walking sounds.

The quantitative results from the experiment showed that the walking sounds were identified with a low identification rate without explicit teaching. This indicated that the previous knowledge of the listeners was not enough to enable good identification results. After the teaching session, the identification rates were considerably better than before teaching. This means that the sound samples themselves contained enough easily distinguishable information to enable correct identification and therefore can be used as auditory icons to represent the person. The average weight of confidence given for the correct answers was greater, and the difference between the weights given for the correct and the incorrect answers was statistically significant after the teaching part. This indicates that the teaching improves also the confidence of the identification.

4.1. Characteristics of Walking Sounds

There are certain issues that may have decreased the identification rates and which should be considered when recording and designing sound for such use as this. One of the most important issues in this sense is that the walking sounds are consistent with those the user already knows. The recorded walking sound samples used in the test differed from the naturally occurring sounds, both in walking, but also in the quality of the sample. The sound samples were recorded in the laboratory environment, where people were asked to walk circle around the microphone. The unnatural situation and knowledge that the walking sounds are recorded may have affected the user's way of walking. Recording in an everyday situation without the walker being aware of the recording could have produced more natural sounds. However, this would technically very challenging. Due to the used microphone and conversion into digital form, the sounds were altered during the recording and some frequencies are emphasized. This has brought up noises that are normally quiet, for example, the squeaking of shoes. The walking samples were also recorded by using just one set of shoes and clothes.

With this test setting it was difficult to find out whether the listeners learned to identify the walking sounds or the characteristics of the sound samples. However, the informal discussions gave cues on the fact that the people used both elements of walking and the characteristics of the sample to recognize the person. They utilized, for example the elements of walking such as rhythm, tempo and heaviness of the step. In the question forms listeners were asked to define the elements of walking sounds they used in the recognition task. The speed, tempo or rate of the walking and sweeping of feet were mentioned in eight cases out of twelve. Shoe sounds, such as squeaking and other noises such as sound of clothes were mentioned seven times. Shoe type and characteristics were mentioned three times, as well as the clacking of the sole and heel. Heaviness of the step appeared three times and tinkling of some chain or some other accessory appeared four times. Pitch of the footstep sound was mentioned two times. Rhythm of walking and strength/volume of the step were also named.

4.2. Two ways of Recognizing Walking Sounds

When people cannot identify the walker directly, they utilize their knowledge and impression about the possible walkers. The listener can classify the sound and restrict the possible identity of the walker. This is a benefit compared to auditory icons that do not have an intuitive connection to the person they represent.

Informal discussions after the second part indicated also that seeing the picture of the person creating the sound helped the listener in the identification. This may have been because the picture gave information on the sources of sound, such as clothing and shoes, and the character of the person.

In the questioning forms and informal discussions occurred two approaches to the identification the walking sounds. One approach is to associate the person into the sound pattern of walking by hearing it. Learning the association happens through visual or aural confirmation perceived together with the walking sound and several repetitions of the sound. The identification itself happens by matching the sound into the learned model.

The other approach is to create the association between the impression of the person and the characteristics of the sound. This is similar to the ecological approach suggested by Gaver [3] (after Gibson [4]): the user is able to recognize the source of the sound, its characteristics and the event creating the sound. The characteristics of walking sound are considered as a manifestation of the persons' characteristics. Rather than identifying the person's walking, the listener recognizes the characteristics of the walker and tries to match these into the knowledge and impression they have on the possible person. For example, a laid-back person is expected to walk loosely and slowly and some other person is known to be limping the other feet. Such characteristics of the walker can also be the shoes or clothing affecting the sound.

This information was applied also in the other way. The listeners applied their knowledge on the elements worn by a possible person, which could affect the walking sound. These kinds of elements were, for example, certain types of heavy boots or jewelry making sounds.

The listener's in the experiment applied this method especially in the case where they were not able to recognize the person directly on the base of the walking sound. The design of walking sounds should be made by keeping these two approaches in mind.

4.3. Ways to Design Better Sound Samples

Supporting the identification by using both approaches can be helped in two ways. By emphasizing certain identifiable, possibly expected elements of the natural sounds preserves the original walking sound, but provides cues for those using the identification of the characteristics of the person. For example, emphasizing the sweep of the other feet or the identifiable noise of a key chain could enhance the identification. The original walking sound could be enhanced with additional auditory elements adding some known and expected, recognizable elements to the sound. For example, adding sound elements such as tinkling of a key chain or sound of some identifiable textile could support forming the impression about the person. To recognize the identifiable and expected characteristics of the walking of a person, people could be asked their description of the person's walking sounds.

Opposite approach to the design of walking sounds would be entirely synthesized walking sounds based on the expectations and knowledge of the listeners. Recognizable sound elements such as sweeps, knocks, trouser legs and squeaking of shoes can be used. The other method could be the use of synthesizers. The method introduced by Cook [5] suggests capturing the envelope of the walking sound samples and using the envelope to synthesize artificial walking sounds. The method preserves the characteristics of the original walking sound, but enables modifying the parameters of the sound and this way, changing the perceived characteristics of the sound source.

5. A SYSTEM FOR SUPPORTING AWARENESS

The presented work is a part of the ubiquitous computing system Doorman. The system enhances group work in office environments, one of its aims being to provide information between group members. The use of walking sounds to support awareness between group members is one of the functionalities of the system. The Doorman system is presented in more detail in [6].

There are two ways to use walking sounds as the part of an application. First, the walking sounds can be played by using public speakers in places where all group members can hear them. If used in this way, they are more like the natural walking sounds. There are problems, however. First, they can be quite obtrusive, since they should be played loud in order to be audible to all group members. Furthermore, itís not possible to personalize or turn off the sounds. It is also possible that the users confuse real and recorded walking sounds causing the recognitions done by applying the spatial context be incorrect.

The other possibility is to use the walking sounds locally for those group members, who need them. In this way, walking sounds could be personalized for each user. The walking sounds could be less obtrusive, and people can always turn them off. The learning could be also supported locally, so for example we could add speech to describe each sound, and when the user has learned the sounds, he/she could turn the speech off. Users could also personalize their sounds, for example to add additional sounds to the basic walking sounds.

Technically, the Doorman system uses the Jaspis architecture [7] to distribute walking sounds for different group members and handle the application logic. We are using loudspeakers which have narrow sound field, which means that we can position the sounds to individual group members even if they are working in the same room.

6. CONCLUSIONS

Using the walking sounds as auditory icons that inform the presence of co-workers include several attractive benefits. The walking sounds as the natural part of the sound ecology are non-obtrusive. Yet they naturally convey the information we are trying to provide to the user. However, our test showed that without any training, simple recorded walking sounds were too hard to recognize. A teaching session improved the recognition rates significantly but such training is not a very desirable requirement. The careful design of walking sound based auditory icons could however help this problem. By supporting the impression people have about each other could make the sounds more distinguishable without losing benefits of prior knowledge.

7. REFERENCES

[1] Li, X.F., Logan, R.J., Pastore, R.E. Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America*, vol. 90 no. 6, 1991, 3036-3049.

[2] Gaver, W.W., Moran, T., MacLean, A., Lˆvstrand, L., Dourish, P., Carter, K., Buxton, W. Realizing a video environment: Europarc's RAVE system. *Conference Proceedings of Conference on Human Factors in Computing Systems (CHI'92)*, 1992, 27-35.

[3] Gaver, W.W. Using and Creating Auditory Icons. In: G. Kramer, G. (ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Santa Fe Institute Studies in the Sciences of Complexity, Addison Wesley, 1994, 417-446.

[4] Gibson, J.J. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

[5] Cook, P.R. Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds. *Proceedings of the AES 22nd International Conference*, 2002, 73-78.

[6] Mäkelä, K., Salonen, E.-P., Turunen, M., Hakulinen, J., Raisamo, R. Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman. *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, Verona, December 14-15, 2001, 115 - 119.

[7] Turunen, M., Hakulinen, J. Jaspis - A Framework for Multilingual Adaptive Speech Applications. *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, 2000.